CIRJE-F-670

A General Asymptotic Theory for Time Series Models

Shiqing Ling
Hong Kong University of Science and Technology

Michael McAleer
Erasmus University Rotterdam
and Tinbergen Institute
and CIRJE, Faculty of Economics, University of Tokyo

September 2009

A General Asymptotic Theory for Time Series Models *

Shiqing Ling
Department of Mathematics
Hong Kong University of Science and Technology

Michael McAleer
Econometric Institute, Erasmus School of Economics
Erasmus University Rotterdam
and Tinbergen Institute, The Netherlands
and Center for International Research on the Japanese Economy (CIRJE)
Faculty of Economics, University of Tokyo

Revised: September 2009

Abstract

This paper develops a general asymptotic theory for the estimation of strictly stationary and ergodic time series models. Under simple conditions that are straightforward to check, we establish the strong consistency, the rate of strong convergence and the asymptotic normality of a general class of estimators that includes LSE, MLE, and some M-type estimators. As an application, we verify the assumptions for the long-memory fractional ARIMA model. Other examples include the GARCH(1,1) model, random coefficient AR(1) model and the threshold MA(1) model.

Key words and phrases: Asymptotic normality, estimation, rate of strong convergence, strong consistency, time series models.

1 Introduction

The three main results that can be used for the asymptotic theory of the estimators in time series models are Basawa, Feign and Heyde (1976), Amemiya (1985) and

*The first author thanks the Hong Kong Research Grants Commission for Grant #HKUST4765/03H. The second author is most grateful for the financial support of the Australian Research Council and the National Science Council, Taiwan.

Tjϕstheim (1986). While not specific to time series models, Basawa et al. (1976) and Amemiya (1985) provide the condition for the weak consistency of the estimated parameters. For asymptotic normality, the result in Basawa et al. (1976) requires the expectation of the third derivatives of the objective function (OF). The condition in Amemiya (1985) does not give a specific method for the convergence of the sample information matrix to prove asymptotic normality.

The result in Tjϕstheim (1986) holds for strictly stationary and ergodic time series models, and implies that there exists a sequence of strongly consistent estimators to maximize the OF. However, this sequence of estimators may not be the global maximizer of the OF. His result also requires the expectation of the third derivatives of the OF, but the third derivatives of the OF can be extremely complicated in some models, as in the case of the likelihood function for ARMA-GARCH models. Jeantheau (1998) also gives the condition for strong consistency of the maximum likelihood estimator (MLE) for a class of GARCH models. However, the results in each of these papers do not discuss the initial value problem, which needs to be addressed for each individual model. As many time series models have been developed in the last two decades, a unified and simple asymptotic theory of estimation for time series models should have wide applicability.

This paper establishes a general asymptotic theory for the estimation of strictly stationary and ergodic time series models. The estimators, including LSE, MLE, and some M-type estimators (except for LAD estimator), among others, are the global maximizers of the respective OFs. We establish the strong consistency, the rate of strong convergence and the asymptotic normality of the estimated parameters. The rate of strong convergence of the estimated parameters has not previously appeared in the literature in a general setting. The conditions, including the initial conditions, are simple and easy to check, and third derivatives are not required. Some related references are Huber(1967) and Pfanzagl (1969).

This paper proceeds as follows. Section 2 presents the model and the main results. Section 3 examines the long-memory FARIMA model. The proofs are given in Sections 4-5.

2 Model and Main Results

Assume that the real $p \times 1$ vector time series $\{y_t : t = 0, \pm 1, \cdots\}$ is $\mathcal{F}_t-$ measurable, strictly stationary and ergodic, and its conditional distribution is given by

$$(2.1) \qquad y_t | \mathcal{F}_{t-1} \quad \sim \quad G(\theta, Y_{t-1}),$$

where $\mathcal{F}_t$ is the σ-field generated by $\{y_t, y_{t-1}, \cdots\}$, $Y_t = (y_t, \cdots, y_{t-p+1})$ or $Y_t = (y_t, y_{t-1}, \cdots)$, and θ is an $m \times 1$ unknown parameter vector. The structure of the time series $\{y_t\}$ is characterized by the distribution G and the parameter θ. We assume that the parameter space Θ is a compact subset of R^m, and the true value θ_0 of θ is an interior point in Θ. We use the following OF with the initial value Y_0 to estimate θ_0:

$$L_n(\theta) = \sum_{t=1}^{n} l_t(\theta),$$

where $l_t(\theta) = l(Y_t, \theta)$ is a measurable function with respect to Y_t and is almost surely (a.s.) and continuously twice differentiable in terms of θ.

Denote $D_t(\theta) = \partial l_t(\theta)/\partial\theta$, $P_t(\theta) = -\partial^2 l_t(\theta)/\partial\theta\partial\theta'$, $\Sigma = E[P_t(\theta_0)]$ and $\Omega = E[D_t(\theta_0)D_t'(\theta_0)]$. Let $V_0(\eta) = \{\theta : \|\theta - \theta_0\| < \eta\}$. The following assumption is made:

Assumption 2.1.

(i) $E \sup_{\theta\in\Theta}[l_t(\theta)] < \infty$, *and* $E[l_t(\theta)]$ *has a unique maximizer at* θ_0*;*

(ii) $D_t(\theta_0)$ *is a martingale difference in terms of* $\mathcal{F}_t$ *with* $0 < \Omega < \infty$*;*

(iii) $\Sigma > 0$ *and* $E \sup_{\theta\in V_0(\eta)} \|P_t(\theta)\| < \infty$ *for some* $\eta > 0$.

When model (2.1) reduces to the class of models: $y_t = f(\theta, Y_{t-1}) + \eta_t\sqrt{h(\theta, Y_{t-1})}$ with $\{\eta_t\}$ being i.i.d. with mean zero and $E\eta_t^2 = 1$, and the QMLE is used, we have

$$l_t(\theta) = -\frac{1}{2}\Big[\log h_t(\theta) + \frac{\varepsilon_t^2(\theta)}{h_t(\theta)}\Big],$$

where $f(\lambda, Y_t)$ and $h(\lambda, Y_t) > 0$ are measurable functions in terms of Y_t. In this case, Assumption 2.1 (i) is ensured by the conditions: (a) $E\sup_{\theta\in\Theta}[\varepsilon_t^2(\theta)/\ h_t(\theta)] < \infty$ and $E\sup_{\theta\in\Theta}\log h_t(\theta) < \infty$, and (b) $[\varepsilon_t(\theta), h_t(\theta)] = [\varepsilon_t(\theta_0), h_t(\theta_0)]$ a.s. if and only if $\theta = \theta_0$. See Jeantheau (1998) and Ling and McAleer (2003).

When the dimension of the initial value Y_0 is infinite, it need to be replaced by some constant $\widetilde{Y}_0$. We denote $l_t(\theta)$ with the initial value $\widetilde{Y}_0$ by $\widetilde{l}_t(\theta)$. Similarly define $\widetilde{D}_t(\theta)$ and $\widetilde{P}_t(\theta)$. The initial condition is given as follows.

Assumption 2.2. *For some $\nu > 0$, it follows that*

(i) $E\sup_{\Theta}|l_t(\theta) - \widetilde{l}_t(\theta)| = O(\frac{1}{t^\nu})$;

(ii) $E\|D_t(\theta_0) - \widetilde{D}_t(\theta_0)\| = O(\frac{1}{t^{1/2+\nu}})$ *and* (iii) $E\sup_{\Theta}\|P_t(\theta) - \widetilde{P}_t(\theta)\| = O(\frac{1}{t^\nu})$.

The decay rates in Assumption 2.2 are very low and are satisfied by most of time series models. For long memory time series, Assumption 2.2(ii) can be replaced by:

Assumption 2.2*(ii'). For any $\epsilon > 0$,*

$$\lim_{l\to\infty} P(\max_{l\leq n<\infty}\frac{1}{\sqrt{n}}\|\sum_{t=1}^{n}[D_t(\theta_0) - \widetilde{D}_t(\theta_0)]\| > \epsilon) = 0.$$

The corresponding OF is modified as

$$\widetilde{L}_n(\theta) = \sum_{t=1}^{n}\widetilde{l}_t(\theta).$$

When the dimension of Y_0 is finite, Assumption 2.2 is redundant. In what follows, $\longrightarrow_{\mathcal{L}}$ denotes convergence in distribution. We now state our main result as follows:

Theorem 2.1 *Let $\hat{\theta}_n = argmax_{\Theta}\widetilde{L}(\theta)$.*

(a) If Assumptions 2.1(i) and 2.2(i) hold, then $\hat{\theta}_n \to \theta_0$ a.s..

(b) Furthermore, if Assumptions 2.1(ii)-(iii) and 2.2(ii)-(iii) hold, then

$$\hat{\theta}_n = \theta_0 + O\Big[\Big(\frac{\log\log n}{n}\Big)^{1/2}\Big] \ a.s. \ and \ \sqrt{n}(\hat{\theta}_n - \theta_0) \longrightarrow_{\mathcal{L}} N(0, \Sigma^{-1}\Omega\Sigma^{-1}).$$

Remark 2.1. For (a), we only need $l_t(\theta)$ to be continuous in terms of θ a.s., while twice differentiability is redundant. $\hat{\theta}_n$ can be LSE, MLE, and some M-type estimators, among others. Many nonlinear time series models in Tong (1990) satisfy

Assumptions 2.1-2.2, such as TAR, bilinear ARMA, GARCH and random coefficient AR models. The compactness of Θ is not a serious restriction in practice since the true value θ_0 is an interior point in the parameter space and we can always get a compact Θ to include it. Compared with the assumptions mentioned in Section 1, our assumptions are simple, clear and easy to check in practice.

Σ and Ω can be estimated, respectively, by

$$\hat{\Sigma}_n = \frac{1}{n}\sum_{t=1}^{n} \widetilde{P}_t(\hat{\theta}_n) \ \ and \ \ \hat{\Omega}_n = \frac{1}{n}\sum_{t=1}^{n} \widetilde{D}_t(\hat{\theta}_n)\widetilde{D}_t'(\hat{\theta}_n).$$

By Lemma 4.2, $\hat{\Sigma}_n = \Sigma + o(1)$ a.s.. Since $\sup_\Theta \|P_t(\theta)\|^{1/2}$ is strictly stationary and has a finite variance, we know that $\max_{1\le t\le n}\sup_\Theta \|P_t(\theta)\|^{1/2}/\sqrt{n} = o_p(1)$. By Taylor's expansion, we have $\|D_t(\hat{\theta}_n) - D_t(\theta_0)\|/\sqrt{n} \le [\max_\Theta \|P_t(\theta)\|/n]\|\sqrt{n}(\hat{\theta}_n - \theta_0)\| = o_p(1)$ uniformly in t. Furthermore, by Assumption 2.2(ii), we know that $\|\widetilde{D}_t(\hat{\theta}_n) - D_t(\theta_0)\|/\sqrt{n} = o_p(1)$ uniformly in t. By Taylor's expansion and Lemma 4.2, we have $\sum_{t=1}^n \widetilde{D}_t(\hat{\theta}_n)/\sqrt{n} = -n^{-1}\sum_{t=1}^n \widetilde{P}_t(\hat{\theta}_n^*)\sqrt{n}(\hat{\theta}_n - \theta_0) = O_p(1)$. Thus, $n^{-1}\sum_{t=1}^n[\widetilde{D}_t(\hat{\theta}_n) - D_t(\theta_0)]\widetilde{D}_t'(\hat{\theta}_n)] = o_p(1)$. Finally, by the ergodic theorem, we can see that $\hat{\Omega}_n = \Omega + o_p(1)$. Thus, under Assumptions 2.1-2.2, $\hat{\Sigma}_n$ and $\hat{\Omega}_n$ are consistent estimators of Σ and Ω, respectively.

Example 2.1. Consider the GARCH(1, 1) model:

$$y_t = \eta_t\sqrt{h_t} \text{ and } h_t = \alpha_0 + \alpha y_{t-1}^2 + \beta h_{t-1},$$

where $\eta_t \sim$i.i.d. N(0,1), $\alpha_0 > 0$, $\alpha > 0$ and $\beta > 0$. Assume that $E\ln(\beta + \alpha\eta_t^2) < 0$. Let $\theta = (\alpha_0, \alpha, \beta)'$. When using MLE to estimate θ_0, we take

$$l_t(\theta) = -\log h_t(\theta) - \frac{y_t^2}{h_t(\theta)},$$

where $h_t(\theta) = \alpha_0 + \alpha y_{t-1}^2 + \beta h_{t-1}(\theta)$, $t = 1, \cdots, n$ and $y_0 = 0$ and $h_0 = 1$. From the proof of Theorem 2.1 in Francq and Zakoan (2004), we see that Assumptions 2.1(i) and 2.2(i) hold. From the proof of Theorem 2.2 in Francq and Zakoian (2004), we know that Assumptions 2.1(ii)-(iii) and 2.2(ii)-(iii) hold, see also Lee and Hansen (1994) and Ling and McAleer (2003). We do not need to study the third derivative of $l_t(\theta)$ as done in these papers.

Example 2.2. Consider the random coefficient AR(1) model:

$$y_t = (\phi + \psi_t)y_{t-1} + \varepsilon_t,$$

where $\{\psi_t\}$ and $\{\varepsilon_t\}$ are i.i.d. sequences with zero mean and variance $\alpha > 0$ and $\sigma^2 > 0$, respectively, and they are mutually independent. Assume $E \ln |\phi + \psi_t| < 0$ and let $\theta = (\phi, \alpha, \sigma^2)'$. When we use QMLE to estimate θ_0, we take

$$l_t(\theta) = -\frac{1}{2}\log(\sigma^2 + \alpha y_{t-1}^2) - \frac{(y_t - \phi y_{t-1})^2}{2(\sigma^2 + \alpha y_{t-1}^2)}.$$

By exactly following the proof of Lemmas A.1, A.2 and A.3(i) in Ling (2004), we can show that Assumption 2.1 holds. Assumption 2.2 hold automatically.

Example 2.3. Consider the first order threshold MA (TMA (1)) model:

$$y_t = [\phi + \psi I(y_{t-1} \le r)]\varepsilon_{t-1} + \varepsilon_t,$$

where $\{\varepsilon_t\}$ is a sequence of i.i.d random variables, with mean zero, variance $0 < \sigma^2 < \infty$ and a density function f. Assume that $|\phi| < 1$, $|\phi + \psi| < 1$ and $|\psi| \sup_x |xf(x)| < 1$. This assumption ensures that the TMA(1) model is strictly stationary and ergodic, and invertible, see Ling, Tong and Li (2007). Let $\theta = (\phi, \psi)$. When using the CLSE to estimate θ_0, we take

$$l_t(\theta) = -\varepsilon_t^2(\theta),$$

where $\varepsilon_t(\theta) = y_t - [\phi + \psi I(y_{t-1} \le r)]\varepsilon_{t-1}(\theta)$, $t = 1, \cdots, n$ and $\varepsilon_t(\theta) = 0$ as $t \le 0$. Furthermore, assume that the delay parameter r is known and $E\varepsilon_t^4 < \infty$. By the very minor modification of Lemmas 6.1-6.5 in Ling and Tong (2005), we can show that Assumption 2.1 holds. Similarly, a minor modification of Lemma 6.6 in Ling and Tong (2005) shows that Assumption 2.2 holds. We should mention that this is a new result for the TMA(1) model. When r is unknown, the asymptotic theory on the TMA model remains open.

3 Application to Long Memory FARIMA Models

The process $\{y_t\}$ is said to follow the long memory (LM)-ARFIMA model if

$$\phi(B)(1-B)^d y_t = \psi(B)\varepsilon_t, \tag{3.1}$$

where $\phi(B) = 1 - \sum_{i=1}^{p} \phi_i B^i$, $\psi(B) = 1 + \sum_{i=1}^{q} \psi_i B^i$, $(1-B)^d = \sum_{k=0}^{\infty} a_k B^k$ with $a_k = (k-d-1)!/k!(-d-1)!$, B is the backward-shift operator, and $\{\varepsilon_t\}$ is i.i.d. with zero mean and variance $0 < \sigma^2 < \infty$. $\theta = (d, \phi_1, \cdots, \phi_p, \psi_1, \cdots, \psi_q)'$ and its true value is θ_0. We assume that the parameter space Θ is a compact subset of R^{p+q+1}, θ_0 is an interior point in Θ, and the following assumption holds.

Assumption 3.1. *$d \in (0, 1/2)$, all the roots of $\phi(B)$ and $\psi(B)$ are outside the unit circle, $\phi_p \neq 0$, $\psi_q \neq 0$, and $\phi(B)$ and $\psi(B)$ have no common root.*

Given $\{y_1, \cdots, y_n\}$, we consider the conditional LSE of θ_0, which is defined as $\hat{\theta}_n = \operatorname{argmin} \sum_{t=1}^{n} \tilde{\varepsilon}_t^2(\theta)$, where $\tilde{\varepsilon}_t(\theta)$ is $\varepsilon_t(\theta) = \psi^{-1}(B)\phi(B)(1-B)^d y_t$, with initial value $\widetilde{Y}_0$. We have the following results:

Theorem 3.1. *If Assumption 3.1 holds, then*

$$\begin{aligned}
&(a) \qquad \hat{\theta}_n = \theta_0 + O\Big[\Big(\frac{\log\log n}{n}\Big)^{1/2}\Big] \ a.s., \\
&(b) \qquad \sqrt{n}(\hat{\theta}_n - \theta_0) \longrightarrow_{\mathcal{L}} N\Big(0, \sigma^2 E^{-1}\Big[\frac{\partial \varepsilon_t(\theta_0)}{\partial \theta}\frac{\partial \varepsilon_t(\theta_0)}{\partial \theta'}\Big]\Big).
\end{aligned}$$

Remark 3.1. Model (3.1) has the long-memory property and has been widely applied in hydrology and economics. Some related references are Granger and Joyeux (1980), Hosking (1981), Li and McLeod (1986), Robinson (1994) and Beran (1995), among others. When ε_t follows the GARCH model, model (3.1) was studied by Baillie (1996), Ling and Li (1997) and Ling (2003). However, the paper is the first to provide the rate of strong convergence of $\hat{\theta}_n$, as in (a). From the proof in Section 5, we can see that the initial condition is crucial in this development.

4 Proof of Theorem 2.1

Lemma 4.1. *If Assumptions 2.1(i)-2.2(i) hold, then for any $\eta > 0$,*

$$\lim_{l\to\infty} P\Big(\max_{l\le n<\infty} \sup_{\|\theta-\theta_0\|\ge\eta} \sum_{t=1}^{n}[\tilde{l}_t(\theta) - \tilde{l}_t(\theta_0)] \ge 0\Big) = 0.$$

Proof. Let $V_{\tilde{\eta}} = \{\tilde{\theta} : \|\tilde{\theta} - \theta\| \le \tilde{\eta}\}$ and $X_t(\tilde{\eta}) = \sup_{\theta\in\Theta}\sup_{V_{\tilde{\eta}}} |l_t(\tilde{\theta}) - l_t(\theta)|$. By Assumption 2.1(i), $EX_t(\tilde{\eta}) \to 0$ as $\tilde{\eta} \to 0$. Thus, for any $\epsilon > 0$, there is $\tilde{\eta} > 0$ such that $EX_t(\tilde{\eta}) < \epsilon/2$. Since $X_t(\tilde{\eta})$ is strictly stationary and ergodic, by Lemma 1 in Chow and Teicher (1978, p.66) and the ergodic theorem, for any $\epsilon_1 > 0$, we have

$$P(\max_{l\le n<\infty} \frac{1}{n}\Big|\sum_{t=1}^{n}[X_t(\tilde{\eta}) - EX_t(\tilde{\eta})]\Big| \ge \frac{\epsilon}{2}) < \epsilon_1,$$

as l is large enough. Thus, for any $\epsilon, \epsilon_1 > 0$, there exists a constant $\tilde{\eta} > 0$ such that

$$P(\max_{l\le n<\infty} \frac{1}{n}\sum_{t=1}^{n} X_t(\tilde{\eta}) \ge \epsilon) \le P(\max_{l\le n<\infty} \frac{1}{n}\Big|\sum_{t=1}^{n}[X_t(\tilde{\eta}) - EX_t(\tilde{\eta})]\Big| \ge \frac{\epsilon}{2}) < \epsilon_1. \tag{4.1}$$

By Assumption 2.1(i) and the ergodic theorem, for each $\theta \in \Theta$ and any $\epsilon > 0$,

$$\lim_{l\to\infty} P\Big(\max_{l\le n<\infty} \Big|\frac{1}{n}\sum_{t=1}^{n}[l_t(\theta) - El_t(\theta)]\Big| \ge \epsilon\Big) = 0. \tag{4.2}$$

Since Θ is compact, we can choose a collection of balls of radius $\Delta > 0$ covering Θ, and the number of such balls is a finite integer N. In the ith ball, we take a point ξ_i and denote this ball by $V(\xi_i)$. For any $\epsilon > 0$,

$$\begin{aligned}
&P(\max_{l\le n<\infty} \frac{1}{n}\sup_{\Theta}\Big|\sum_{t=1}^{n}[l_t(\theta) - El_t(\theta)]\Big| \ge \epsilon) \\
&\le P\Big(\max_{1\le j\le N}\sup_{\theta\in V(\xi_j)}\max_{l\le n<\infty}\Big|\frac{1}{n}\sum_{t=1}^{n}[l_t(\theta) - l_t(\xi_j)]\Big| \ge \frac{\epsilon}{3}\Big) \\
&\quad + P\Big(\max_{1\le j\le N}\sup_{\theta\in V(\xi_j)}\Big|E[l_t(\theta) - l_t(\xi_j)]\Big| \ge \frac{\epsilon}{3}\Big) \\
&\quad + P\Big(\max_{1\le j\le N}\max_{l\le n<\infty}\Big|\frac{1}{n}\sum_{t=1}^{n}[l_t(\xi_j) - El_t(\xi_j)]\Big| \ge \frac{\epsilon}{3}\Big) \\
&\le P\Big(\sup_{\xi_i\in\Theta}\sup_{\theta\in V(\xi_j)}\max_{l\le n<\infty}\Big|\frac{1}{n}\sum_{t=1}^{n}[l_t(\theta) - l_t(\xi_j)]\Big| \ge \frac{\epsilon}{3}\Big) \\
&\quad + \sum_{j=1}^{N} P\Big(\sup_{\theta\in V(\xi_j)}\Big|E[l_t(\theta) - l_t(\xi_j)]\Big| \ge \frac{\epsilon}{3}\Big) \\
&\quad + \sum_{j=1}^{N} P\Big(\max_{l\le n<\infty}\Big|\frac{1}{n}\sum_{t=1}^{n}[l_t(\xi_j) - El_t(\xi_j)]\Big| \ge \frac{\epsilon}{3}\Big) < \epsilon,
\end{aligned} \tag{4.3}$$

as l is large enough and Δ is small enough, where the last inequality holds by (4.1), (4.2) and the uniform continuity of $El_t(\theta)$.

Since $E[l_t(\theta)]$ has a unique maximum at θ_0, Θ is compact, and $El_t(\theta)$ is continuous, there exists a constant $c > 0$, such that

$$\max_{\|\theta-\theta_0\|>\eta} E[l_t(\theta) - l_t(\theta_0)] \le -c, \tag{4.4}$$

for any $\eta > 0$. By (4.3)-(4.4), it follows that

$$\begin{aligned}
&P\Big(\max_{l\le n<\infty} \sup_{\|\theta-\theta_0\|\ge\eta} \{\sum_{t=1}^n [l_t(\theta) - l_t(\theta_0)] + \frac{cn}{2}\} > 0\Big) \\
&= P\Big(\max_{l\le n<\infty} \sup_{\|\theta-\theta_0\|\ge\eta} \{\sum_{t=1}^n [l_t(\theta) - El_t(\theta)] \\
&\quad - \sum_{t=1}^n [l_t(\theta_0) - El_t(\theta_0)] + n[El_t(\theta) - El_t(\theta_0)] + \frac{cn}{2}\} > 0\Big) \\
&\le P\Big(\max_{l\le n<\infty} \sup_{\Theta}\{2\Big|\sum_{t=1}^n [l_t(\theta) - El_t(\theta)]\Big| - cn + \frac{cn}{2}\} > 0\Big) \\
&\le P\Big(\max_{l\le n<\infty} \sup_{\Theta}\{\Big|\frac{1}{n}\sum_{t=1}^n [l_t(\theta) - El_t(\theta)]\Big|\} > \frac{c}{4}\Big) \to 0,
\end{aligned} \tag{4.5}$$

as $l \to \infty$. By Assumption 2.2(i) and Markov's inequality, it follows that

$$\begin{aligned}
&\lim_{l\to\infty} P(\max_{l\le n<\infty} \frac{1}{n}\sum_{t=1}^n \sup_{\Theta} |\tilde{l}_t(\theta) - l_t(\theta)| > \epsilon) \\
&\le \lim_{l\to\infty} P(\sum_{t=1}^\infty \frac{1}{t^{1-0.5\nu}} \sup_{\Theta} |\tilde{l}_t(\theta) - l_t(\theta)| > \epsilon l^{0.5\nu}) \\
&\le \lim_{l\to\infty} \frac{1}{\epsilon l^{0.5\nu}} \sum_{t=1}^\infty \frac{1}{t^{1-0.5\nu}} E \sup_{\Theta} |\tilde{l}_t(\theta) - l_t(\theta)| = \lim_{l\to\infty} \frac{O(1)}{\epsilon l^{0.5\nu}} \sum_{t=1}^\infty \frac{1}{t^{1+0.5\nu}} = 0,
\end{aligned} \tag{4.6}$$

for any $\epsilon > 0$. By (4.5) and (4.6), we have

$$\begin{aligned}
&P\Big(\max_{l\le n<\infty} \sup_{\|\theta-\theta_0\|\ge\eta} \sum_{t=1}^n [\tilde{l}_t(\theta) - \tilde{l}_t(\theta_0)] + \frac{cn}{4} > 0\Big) \\
&\le P\Big(\max_{l\le n<\infty} \sup_{\|\theta-\theta_0\|\ge\eta} \sum_{t=1}^n [l_t(\theta) - l_t(\theta_0)] + 2\max_{l\le n<\infty} \sum_{t=1}^n \sup_{\Theta} |\tilde{l}_t(\theta) - l_t(\theta)| + \frac{cn}{4} > 0\Big) \\
&\le P\Big(\max_{l\le n<\infty} \sup_{\|\theta-\theta_0\|\ge\eta} \sum_{t=1}^n [l_t(\theta) - l_t(\theta_0)] + \frac{cn}{2} > 0\Big) \\
&\quad + P(\max_{l\le n<\infty} \frac{1}{n}\sum_{t=1}^n \sup_{\Theta} |\tilde{l}_t(\theta) - l_t(\theta)| > \frac{c}{8}) \to 0,
\end{aligned}$$

as $l \to \infty$. By this equation, we can see that the conclusion holds. $\square$

Lemma 4.2. *If Assumptions 2.1(iii)-2.2(iii) hold, then for any* $\epsilon > 0$,

$$\lim_{l\to\infty} P(\max_{l\le n<\infty}\max_{V_0(\eta)}\frac{1}{n}\Big\|\sum_{t=1}^{n}[\widetilde{P}_t(\theta)-\Sigma]\Big\|\ge\epsilon)=0.$$

Proof. Let $X_t(\eta)=\sup_{V_0(\eta)}\|P_t(\theta)-P_t(\theta_0)\|$. By Assumption 2.1(iii), as η is small enough, $EX_t(\eta)<\epsilon/4$. Since $\{X_t(\eta)\}$ is strictly stationary and ergodic, by Lemma 1 in Chow and Teicher (1978, p.66) and the ergodic theorem, we have

$$\lim_{l\to\infty} P(\max_{l\le n<\infty}\frac{1}{n}\sum_{t=1}^{n}X_t(\eta)\ge\epsilon)\le\lim_{l\to\infty} P(\max_{l\le n<\infty}\frac{1}{n}\|\sum_{t=1}^{n}(X_t(\eta)-EX_t(\eta))\|\ge\frac{\epsilon}{2})=0,$$

for any $\epsilon>0$. By Assumption 2.1(iii) and the ergodic theorem, it follows that

$$\lim_{l\to\infty} P(\max_{l\le n<\infty}\frac{1}{n}\|\sum_{t=1}^{n}[P_t(\theta_0)-\Sigma]\|\ge\epsilon)=0,$$

for any $\epsilon>0$. By the preceding two equations, there is $\eta>0$ such that

$$\lim_{l\to\infty} P(\max_{l\le n<\infty}\max_{V_0(\eta)}\frac{1}{n}\Big\|\sum_{t=1}^{n}[P_t(\theta)-\Sigma]\Big\|\ge\epsilon)=0. \tag{4.7}$$

Let $\widetilde{X_t}=\sup_{V_0(\eta)}\|\widetilde{P}_t(\theta)-P_t(\theta)\|$. By Assumption 2.2(iii), we can show that $P(\max_{l\le n<\infty}\sum_{t=1}^{n}\widetilde{X_t}/n\ge\epsilon)\to 0$ as $l\to\infty$, for any $\epsilon>0$. Note that $\max_{V_0(\eta)}\|\sum_{t=1}^{n}[\tilde{P}_t(\theta)-P_t(\theta)]\|\le\sum_{t=1}^{n}\tilde{X}_t$. Furthermore, by (4.7), the conclusion holds. $\square$

Proof of Theorem 2.1. By Lemma 4.1, for any $\epsilon>0$, we have

$$\begin{aligned}
&\lim_{l\to\infty} P(\max_{l\le n<\infty}\|\hat{\theta}_n-\theta_0\|>\epsilon)\\
&\quad=\lim_{l\to\infty} P\Big\{\max_{l\le n<\infty}\|\hat{\theta}_n-\theta_0\|>\epsilon,\ \max_{l\le n<\infty}\sum_{t=1}^{n}\Big[\tilde{l}_t(\hat{\theta}_n)-\tilde{l}_t(\theta_0)\Big]\ge 0\Big\}\\
&\quad\le\lim_{l\to\infty} P\Big\{\max_{l\le n<\infty}\sup_{\|\theta-\theta_0\|>\epsilon}\sum_{t=1}^{n}\Big[\tilde{l}_t(\theta)-\tilde{l}_t(\theta_0)\Big]\ge 0\Big\}=0.
\end{aligned}$$

Thus, (a) holds. Applying Taylor's expansion to $\partial\tilde{l}_t(\tilde{\theta}_n)/\partial\theta$ and using Lemma 4.2,

$$\hat{\theta}_n-\theta_0=-\Big[\frac{1}{n}\sum_{t=1}^{n}\widetilde{P}_t(\hat{\theta}_n^*)\Big]^{-1}\frac{1}{n}\sum_{t=1}^{n}\widetilde{D}_t(\theta_0)=-[\Sigma+o(1)]^{-1}\frac{1}{n}\sum_{t=1}^{n}\widetilde{D}_t(\theta_0)\ a.s.,$$

where $\hat{\theta}_n^*$ lies between $\hat{\theta}_n$ and θ_0 and $\hat{\theta}_n^*\to\theta_0$ a.s.. By Assumption 2.2(ii) and using a similar method as for (4.6), we can show that

$$\lim_{l\to\infty} P(\max_{l\le n<\infty}\frac{1}{\sqrt{n}}\|\sum_{t=1}^{n}[\widetilde{D}_t(\theta_0)-D_t(\theta_0)]\|\ge\epsilon)=0,$$

for any $\epsilon > 0$. Thus, we have $\hat{\theta}_n - \theta_0 = -[\Sigma + o(1)]^{-1} \sum_{t=1}^{n} D_t(\theta_0)/n + o(n^{-1/2})$ a.s.. By the law of iterated logarithm, we can claim that $\hat{\theta}_n - \theta_0 = O((\log\log n)/n)^{1/2})$ a.s.. By Assumption 2.1(ii) and the central limit theorem, (b) holds. $\square$

5 Proof of Theorem 3.1

Proof. We verify Assumptions 2.1-2.2 with $l_t(\theta) = -\varepsilon_t^2(\theta)$. For simplicity, we only consider the case with $p = q = 0$, while the general case can be similarly verified.

First, Assumption 3.1 ensures that $\{y_t\}$ is strictly stationary and ergodic with $Ey_t^2 < \infty$, and the following expansions hold:

$$y_t = \sum_{i=0}^{\infty} c_{0i}\varepsilon_{t-i} \ \ and \ \ \varepsilon_t(\theta) = (1-B)^d y_t = \sum_{i=0}^{\infty} a_i(\theta) y_{t-i}, \tag{5.1}$$

where $c_{00} = a_0(\theta) = 1$, $c_{0i} = O(i^{-1+d_0})$ and $a_i(\theta) = O(i^{-1-d})$. Since Θ is compact, there are $\underline{d}$ and $\tilde{d}$ such that $0 < \underline{d} \leq d \leq \tilde{d} < 0.5$. Thus, we have $\sup_{\theta\in\Theta} |a_i(\theta)| = O(i^{-1-\underline{d}})$, and hence it follows that

$$\sup_{\Theta} |\varepsilon_t(\theta)| = \sup_{\Theta} |\sum_{i=0}^{\infty} a_i(\theta) y_{t-i}| \leq |y_t| + O(1)\sum_{i=1}^{\infty} \frac{1}{i^{1+\underline{d}}}|y_{t-i}|.$$

By the Cauchy-Schwarz inequality, we have $E\sup_{\Theta} |\varepsilon_t(\theta)|^2 < \infty$. It is not difficult to show that $-E[\varepsilon_t^2(\theta)]$ has a unique maximum on Θ. Thus, Assumption 2.1(i) holds, and

$$D_t(\theta) = -2\varepsilon_t(\theta)\frac{\partial\varepsilon_t(\theta)}{\partial d} \text{ and } P_t(\theta) = 2\Big[\frac{\partial\varepsilon_t(\theta)}{\partial d}\Big]^2 + 2\varepsilon_t(\theta)\frac{\partial^2\varepsilon_t(\theta)}{\partial d^2},$$

where $\partial\varepsilon_t(\theta)/\partial d = \log(1-B)(1-B)^d y_t = \sum_{i=1}^{\infty} a_{1i}(\theta) y_{t-i}$ and $\partial^2\varepsilon_t(\theta)/\partial d^2 = \log^2(1-B)(1-B)^d y_t = \sum_{i=1}^{\infty} a_{2i}(\theta) y_{t-i}$, with $\sup_{\theta\in\Theta} |a_{ji}(\theta)| = O(i^{-1-\underline{d}})$ as $j = 1, 2$. Using these, it is straightforward to show that Assumption 2.1(ii)-(iii) holds.

We next consider Assumption 2.2. For simplicity, let $\widetilde{Y}_0 = (0, 0, \cdots)$. By (5.1),

$$\begin{aligned} & E[\sup_{\Theta} |\varepsilon_t(\theta) - \tilde{\varepsilon}_t(\theta)|]^2 \\ &= E\Big[\sup_{\Theta} |\sum_{i=t}^{\infty} a_i(\theta) y_{t-i}|\Big]^2 \leq CE\Big(\sum_{i=t}^{\infty} \frac{1}{i^{1+\underline{d}}}|y_{t-i}|\Big)^2 = O(t^{-2\underline{d}}). \end{aligned}$$

It is readily shown that $E\sup_\Theta \tilde{\varepsilon}_t^2(\theta)$ is bounded uniformly in t. Thus, by the Cauchy-Schwarz inequality, we have

$$\begin{aligned}
& E\sup_\Theta |\varepsilon_t^2(\theta) - \tilde{\varepsilon}_t^2(\theta)| \\
& \quad \le \{E[\sup_\Theta |\varepsilon_t(\theta) + \tilde{\varepsilon}_t(\theta)|]^2 E[\sup_\Theta |\varepsilon_t(\theta) - \tilde{\varepsilon}_t(\theta)|]^2\}^{1/2} = O(t^{-\underline{d}}),
\end{aligned}$$

so that Assumption 2.2(i) holds. Similarly, we can show that Assumption 2.2(iii) holds.

We now verify Assumption 2.2(ii'). Denote

$$\begin{aligned}
A_t &= \varepsilon_t(\theta_0) - \tilde{\varepsilon}_t(\theta_0) = \sum_{i=t}^{\infty} a_i(\theta_0) y_{t-i}, \\
A_{1t} &= \frac{\partial \varepsilon_t(\theta_0)}{\partial d} - \frac{\partial \tilde{\varepsilon}_t(\theta_0)}{\partial d}, \\
A_{2t} &= \frac{\partial \varepsilon_t(\theta_0)}{\partial d} - v_t = -\sum_{i=t}^{\infty} \frac{1}{i} \varepsilon_{t-i},
\end{aligned}$$

where $v_t = -\sum_{i=1}^{t-1} \varepsilon_{t-i}/i$. We first make the following decomposition:

$$\begin{aligned}
\widetilde{D}_t(\theta_0) - D_t(\theta_0) &= 2\varepsilon_t(\theta_0)\frac{\partial \varepsilon_t(\theta_0)}{\partial d} - 2\tilde{\varepsilon}_t(\theta_0)\frac{\partial \tilde{\varepsilon}_t(\theta_0)}{\partial d} \\
&= 2\varepsilon_t(\theta_0) A_{1t} + 2\frac{\partial \tilde{\varepsilon}_t(\theta_0)}{\partial d} A_t \\
&= 2\varepsilon_t(\theta_0) A_{1t} + 2A_t v_t + 2A_t A_{2t} - 2A_t A_{1t}. \qquad (5.2)
\end{aligned}$$

Since $E(y_t y_{t+r}) = O(|r|^{-1+2d_0})$, we have

$$\begin{aligned}
EA_t^2 &= \sum_{i=t}^{\infty} a_i^2(\theta_0) E y_{t-1-i}^2 + 2\sum_{i=t}^{\infty}\sum_{r=1}^{\infty} a_i(\theta_0) a_{i+r}(\theta_0) E(y_{t-1-i} y_{t-1-i-r}) \\
&= O(1)[\sum_{i=t}^{\infty} \frac{1}{i^{2(1+d_0)}} + 2\sum_{i=t}^{\infty}\sum_{r=1}^{\infty} \frac{1}{i^{1+d_0}(i+r)^{1+d_0} r^{1-2d_0}}] \\
&\le O(1)[\sum_{i=t}^{\infty} \frac{1}{i^{2(1+d_0)}} + 2\sum_{i=t}^{\infty} \frac{1}{i^{1+d_0}} \int_1^{\infty} \frac{1}{(i+x)^{1+d_0} x^{1-2d_0}} dx] \\
&\le O(1)[\sum_{i=t}^{\infty} \frac{1}{i^{2(1+d_0)}} + 2\sum_{i=t}^{\infty} \frac{1}{i^2} \int_0^{\infty} \frac{1}{(1+z)^{1+d_0} z^{1-2d_0}} dz] = O(t^{-1}). \qquad (5.3)
\end{aligned}$$

Using a similar method, we can show that

$$EA_{it}^2 = O(t^{-1}) \text{ and } E(A_t A_{it}) = O(t^{-1}) \text{ as } i = 1, 2. \qquad (5.4)$$

As in the proof of (4.6), using (5.4), we can show that

$$\lim_{l\to\infty} P\Big(\max_{l\le n<\infty}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}|A_tA_{it}|>\epsilon\Big)=0,\ as\ i=1,2. \tag{5.5}$$

We next show that $\sum_{t=1}^{n}\varepsilon_tA_{1t}/\sqrt{n}=o(1)$ a.s. as $n\to\infty$, which is equivalent to

$$\lim_{l\to\infty} P\Big(\max_{l\le n<\infty}\frac{1}{\sqrt{n}}|\sum_{t=1}^{n}\varepsilon_tA_{1t}|>\epsilon\Big)=0. \tag{5.6}$$

By the Kronecker Lemma in Hall and Hedye (1981, p.31), it is sufficient to show that

$$S_k=\sum_{t=1}^{k}\frac{1}{\sqrt{t}}A_{1t}\varepsilon_t\ \ converges\ \ a.s.. \tag{5.7}$$

By (5.4), it follows that

$$E|\sum_{t=s}^{k}\frac{1}{\sqrt{t}}A_{1t}\varepsilon_t|^2=O(1)\Big(\sum_{t=s}^{k}\frac{1}{t^2}\Big)=O(1)\Big(\sum_{t=s}^{k}\frac{1}{t^{1+\nu}}\Big)^{\alpha}, \tag{5.8}$$

for any integer $0<s\le k$ and some $\alpha>1$, where $O(1)$ holds uniformly in k and s. Consider the subsequence $\{S_{2^k}:k=0,1,\cdots\}$. By (5.8), we have

$$E|S_{2^{k+1}}-S_{2^k}|\le O(1)\Big(\sum_{t=2^k+1}^{2^{k+1}}\frac{1}{t^{1+\epsilon}}\Big)\le O\Big(\frac{1}{2^{\epsilon k}}\Big),$$

for some $\epsilon>0$. By this equation and the monotone convergence theorem, we have

$$E\lim_{n\to\infty}\sum_{k=0}^{n}|S_{2^{k+1}}-S_{2^k}|=\lim_{n\to\infty}E\sum_{k=0}^{n}|S_{2^{k+1}}-S_{2^k}|\le O\Big(\sum_{k=0}^{\infty}\frac{1}{2^{\epsilon k}}\Big)<\infty.$$

Thus, $\sum_{k=0}^{n}|S_{2^{k+1}}-S_{2^k}|$ converges a.s. as $n\to\infty$, and hence

$$\lim_{n\to\infty}S_{2^{n+1}}=X_1+\lim_{n\to\infty}\sum_{k=0}^{n}(S_{2^{k+1}}-S_{2^k})\ \ converges\ \ a.s.. \tag{5.9}$$

By (5.8) and Theorem 12.2 in Billingsley (1968), it follows that, for any $\Delta>0$,

$$P\Big(\max_{2^k<n\le 2^{k+1}}|S_n-S_{2^k}|\ge\Delta\Big)\le O(1)\Big(\sum_{t=2^k+1}^{2^{k+1}}\frac{1}{t^{1+\nu}}\Big)^{\alpha}=O\Big(\frac{1}{2^{\epsilon_1 k}}\Big),$$

for some $\epsilon_1>0$. By the Borel-Canteli Lemma, we can claim that

$$\max_{2^k<n\le 2^{k+1}}|S_n-S_{2^k}|\to 0\ a.s.\ as\ k\to\infty. \tag{5.10}$$

By Lemma 2.3.1 in Stout (1974) and (5.9)-(5.10), we know that (5.7) holds.

Note that v_s is independent of A_{1t} as $s \geq 2$, so that

$$\begin{aligned} E|\sum_{t=s}^{k} \frac{1}{\sqrt{t}} A_{1t}\nu_t|^2 &= E|\sum_{t=s}^{k}\sum_{t_1=s}^{k} \frac{1}{\sqrt{t}}\frac{1}{\sqrt{t_1}} E(A_{1t}A_{1t_1})E(\nu_t\nu_{t_1}) \\ &= O(1)\Big(\sum_{t=s}^{k} \frac{1}{t^{1+1/2}}\Big)^2 = O(1)\Big(\sum_{t=s}^{k} \frac{1}{t^{1+\nu}}\Big)^{\alpha}, \end{aligned}$$

for any integer $0 < s \leq k$, where $O(1)$ holds uniformly in k and s. Using this equation and a similar method as for (5.6), we can show that

$$\lim_{l\to\infty} P(\max_{l\leq n<\infty} \frac{1}{\sqrt{n}} |\sum_{t=1}^{n} v_t A_t| > \epsilon) = 0. \tag{5.11}$$

By (5.2), (5.5)-(5.6) and (5.11), we can show that Assumption 2.2(ii') holds. □

REFERENCES

Amemiya, T. (1985), *Advanced Econometrics*, Cambridge, Harvard University Press.

Baillie, R.T. (1996), Long memory processes and fractional integration in econometrics, *Journal of Econometrics* 73, 5-59.

Basawa, I.V., Feign, P.D., and Heyde, C.C. (1976), Asymptotic properties of maximum likelihood estimators for stochastic processes, *Sankhya A*, 259-270.

Beran, J. (1995), Maximum likelihood estimation of the differencing parameter for invertible short and long memory autoregressive integrated moving average models, *Journal of the Royal Statistical Society B* 57, 659-672.

Billingsley, P. (1968), *Convergence of Probability Measures*, New York, Wiley.

Chow, Y.S. and Teicher, H. (1978), *Probability Theory. Independence, Interchangeability, Martingales*, Springer-Verlag, New York and Heidelberg,

Francq, C. and Zakoian, J.M. (2004) Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli* **10**, 605-637.

Granger, C.W.J. and Joyeux, R. (1980), An introduction to long-memory time series models and fractional differences, *Journal of Time Series Analysis* 1, 15-39.

Hall, P. and Heyde, C.C. (1980), *Martingale Limit Theory and Its Applications*, Academic, San Diego.

Hosking, J.R.M. (1981), Fractional differencing, *Biometrika* 68, 165-76.

Huber, P.J. (1967), The behavior of maximum likelihood estimates under nonstandard conditions, in *Proc. 5th Berkeley Symp. Math. Statist. Probab.*, Vol. 1. Univ. California Press, Berkeley, pp. 221-234.

Jeantheau, T. (1998), Strong consistency of estimators for multivariate ARCH models, *Econometric Theory* 14, 70-86.

Lee, S-W. and Hansen, R.E.(1994). Asymptotic theory for GARCH(1,1) quasi-maximum likelihood estimator. *Econometric Theory* **10**, 29-52.

Li, W.K. and McLeod, A.I. (1986), Fractional time series modelling, *Biometrika* 73, 217-21.

Ling, S. (2003), Adaptive estimators and tests of stationary and non-stationary short and long memory ARIMA-GARCH models, *Journal of the American Statistical Association* 98, 955-967

Ling, S. (2004) Estimation and testing of stationarity for double autoregressive models. *Journal of the Royal Statistical Society: Series B* **66**, 63-78.

Ling, S. and Li, W.K. (1997), Fractional ARIMA-GARCH time series models, *Journal of the American Statistical Association* 92, 1184-1194.

Ling, S. and McAleer, M. (2003), Asymptotic theory for a vector ARMA-GARCH model, *Econometric Theory* 19, 280-310.

Ling, S. and Tong, H. (2005) Testing for a linear MA model against threshold MA models. *Annals of Statistics* **33**, 2529-2552..

Ling, S., Tong, H. and Li, D. (2007) Ergodicity and invertibility of threshold MA (1) models. *Bernoulli*, **13**, 161-168.

Pfanzagl, J. (1969), On the measurability and consistency of minimum contrast estimates, *Metrika* 14, 249-272.

Robinson, P.M. (1994), Time series with strong dependence, in C. Sims (ed.), *Advances in Econometrics, Sixth World Congress*, Cambridge, Cambridge University Press.

Stout, W.F.(1974), *Almost Sure Convergence*, Academic Press, New York and London.

Tjϕstheim, D. (1986), Estimation in nonlinear time series models, *Stochastic Processes and Applications* 21, 251-273.

Tong, H. (1990), *Nonlinear Time Series. A Dynamical System Approach*, Clarendon Press and Oxford University Press, New York.